The Patent Office

INVESTOR IN PEOPLE

GB99/3654

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ



**PRIORITY DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

-registration under the Companies Act does not constitute a new legal entity but merely jects the company to certain additional company law rules.

Signed

Dated 24 November 1999

Patents Form 1/77

Act 1[illegible]
(Rule 16)

THE PATENT OFFICE
-7 NOV 1998
RECEIVED BY POST

The Patent Office

09NOV98 E402954-2 D01559
P01/7700 0.00 - 9824357.9

# Request for grant of a patent

*(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)*

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference — WN/KH/WCM.69

2. Patent application number *(The Patent Office will fill in this part)* — 07 NOV 1998 — **9824357.9**

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

University of Wales College of Medicine,
Heath Park,
CARDIFF,
CF4 4XN,

Patents ADP number *(if you know it)*

If the applicant is a corporate body, give the country/state of its incorporation — United Kingdom

FR5400C1

4. Title of the invention

Protein and DNA Coding Therefor

5. Name of your agent *(if you have one)*

Wynne-Jones, Lainé & James

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Morgan Arcade Chambers
33 St Mary Street
Cardiff
CF1 2AB

Patents ADP number *(if you know it)*

1792002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

| Country | Priority application number *(if you know it)* | Date of filing *(day / month / year)* |
|---|---|---|

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

| Number of earlier application | Date of filing *(day / month / year)* |
|---|---|
| - | |

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

*a) any applicant named in part 3 is not an inventor, or*
*b) there is an inventor who is not named as an applicant, or*
*c) any named applicant is a corporate body.*
*See note (d))*

Yes

## Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

| | |
|---|---|
| Continuation sheets of this form | None |
| Description | 28 |
| Claim(s) | 3 |
| Abstract | - |
| Drawing(s) | 16 & 16 |

10. If you are also filing any of the following, state how many against each item. -

| | |
|---|---|
| Priority documents | |
| Translations of priority documents | |
| Statement of inventorship and right to grant of a patent *(Patents Form 7/77)* | |
| Request for preliminary examination and search *(Patents Form 9/77)* | |
| Request for substantive examination *(Patents Form 10/77)* | |
| Any other documents *(please specify)* | |

11. I/We request the grant of a patent on the basis of this application.

Signature Date 6.11.98

Wynne-Jones, Lainé & James

12. Name and daytime telephone number of person to contact in the United Kingdom

Mr. W. Newell
(01222) 229526

**Warning**

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.*

**Notes**

*a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

*b) Write your answers in capital letters using black ink or you may type them.*

*c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

*d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

*e) Once you have filled in the form you must remember to sign and date it.*

*f) For details of the fee and ways to pay please contact the Patent Office.*

# PROTEIN AND DNA CODING THEREFOR

The present invention relates to a protein, capable of bioluminescence, cDNA coding therefor and their uses, *inter alia*, in diagnostics and therapy. In particular, this invention relates to the cloning and sequencing of cDNA coding for pholasin from the bivalve mollusc *Pholas dactylus*.

The term 'bioluminescence' refers to the emission of light resulting from a chemical reaction within, or produced by, a living organism. The essential components to the chemical reaction are: an organic molecule, usually comprising a luciferin; oxygen or one of its metabolites; and an enzyme or luciferase that catalyses the oxidation of the luciferin. The chemiluminescent reaction responsible for bioluminescence may be represented as follows:

Luciferin + $O_2$, or $O^-_2$, or $H_2O_2$, or $^{\cdot}OH$ or $OC1^-$ or $OX^-$ or $^1O_2$ (+ luciferase) $\longrightarrow$ oxyluciferin + light.

Up to three other substances may also be required to generate light or to generate light of the required colour and intensity. These are as follows:

(a) A cation, such as $H^+$, $Ca^{2+}$, $Mg^{2+}$ or a transition metal cation (*eg* $Cu^+/Cu^{2+}$, $Fe^{2+}/Fe^{3+}$, $La^{3+}$ and $V^{3+}$);

(b) A co-factor such as NAD(P)H, FMN or ATP; and/or

(c) A fluor as an energy transfer acceptor.

Five chemical families of luciferin are known:

(a) Aldehydes (found in the freshwater limpet *Latia*, earthworms, and with FMN in bacteria);

(b) Imidazolopyrazines, which are the compounds most commonly responsible for bioluminescence in the sea (found in Sarcomastigophora, Cnidaria, Ctenophora, Annelida, Chaetognatha, some Arthropoda, some Mollusca and some Chordata);

(c) Benzothiazoles (found in beetles such as fireflies and glow-worms);

(d) Linear tetrapyrroles (found in dinoflagellates, euphausiid shrimp and some fish); and

(e) Flavins (found in bacteria, fungi, polychaete worms and some molluscs).

Chemiluminescent reactions involving these luciferins may produce a glow or a flash with an emission of violet, blue, blue-green, green, yellow, orange or red light, or occasionally UV or IR light. The light emission may be linearly or circularly polarised. The luciferin or its product may also be detected and quantified by fluorescence or phosphorescence. As a chemical reaction is directly responsible for the light emission, it does not require exposure to UV, visible or IR light. However, some bioluminescent systems, such as that in the red organ of the deep sea fish *Malacosteus*, exhibit a photo-chemiluminescence, where light can trigger or enhance the chemiluminescent reaction. [Reference is directed to Chemiluminescence: Principles and Applications in Biology and Medicine, A K Campbell (1988), Horwood/VCH Chichester, Weinheim.]

In the case of some bioluminescent proteins, the luciferin is so tightly or covalently bound to the protein molecule that it does not diffuse away into the surrounding fluid as a result of the chemiluminescent reaction. In this case, the protein-luciferin complex is known as a photoprotein; and the protein itself is referred to as an apophotoprotein. Some bioluminescent proteins are proteins whose light emission or radiation depends on or may be altered by oxygen or one of its metabolites; these bioluminescent proteins are hereinafter referred to as 'bioluminescent oxidative indicator proteins' (BOIPs).BOIPS may thus be photoproteins or luciferin-luciferase systems.

BOIPs, therefore, may be used to detect and quantify oxygen or one of its metabolites in individual cells, defined compartments of living cells such as the nucleus, whole organs and organisms - both animals and plants, including microbes such as viruses and bacteria and protozoa - as well as substances of biological interest such as substrates, metabolites, vitamins, drugs, intra- and extra-cellular signals, enzymes, antigens, antibodies and nucleic acids. Heretofore, it has only been known to employ native BOIPs extracellularly.

The present invention therefore relates to a method for the detection and/or measurement of oxygen or one of its metabolites in live cells (intracellular), which method comprises

providing a BOIP, such as native or chemically- (or genetically-) modified BOIP or a 'rainbow protein' based on such a BOIP, intracellularly and thereafter detecting and/or quantifying light emission therefrom and/or changes in colour, intensity and/or polarisation of emission(s) therefrom.

Furthermore, it has now been found that, by sequencing the BOIP and identifying the cDNA encoding therefor, the recombinant BOIP can also be used in such a method, or chemically- or genetically-modified recombinant BOIP, or a 'rainbow protein' based on such a BOIP. For example, the bivalve mollusc *Pholas dactylus* has been shown to comprise a native photoprotein, which interacts with a luciferase, when they are secreted together by the mollusc to produce light when $O_2$ or one of its metabolites is present. References to the Purification and Properties of *Pholas Dactylus* Luciferin and Luciferase can be found by Michelson in Methods in Enzymology LVII 385-406 (1978). References to detection of activation of neutrophils by detection of superoxide anion can be found by Roberts in Anal Biochem 160 139-148 (1987) and by Müller *et al* in J Biolum Chemilum 3 105-113 (1989). The native photoprotein (known as pholasin) is made up of a glycosylated apoprotein (34kDa) with a small organic molecule, the luciferin, tightly bound to it. This luciferin (whose structure is unknown - Müller and Campbell in J Biolum Chemilum 5 25-30 (1990)) can be extracted from the protein moiety - the apopholasin - or from the organism by a standard treatment, such as mild acid. The pholasin may be collected from live molluscs found in sedimentary rocks at low water along the south coast of England from Plymouth to Folkestone and also along the French channel coast and in the Mediterranean. Further details may be obtained from marine fauna and the references cited herein.

We have surprisingly found that pholasin can generate light even without the presence of the corresponding luciferase by addition of oxygen metabolites such as $O_2^-$, $H_2O_2$, OCl- or other oxyhalide anions, or organic peroxides, and certain organic solvents such as dimethyl sulphide (DMSO) or dimethyl formamide (DMF).

We have now identified the cDNA encoding for the (non-glycosylated) apoprotein of pholasin, which may also be called 'apopholasin'. Accordingly, the present invention therefore further provides an isolated, purified or recombinant nucleic acid sequence

comprising:

(a) The apophotoprotein of pholasin (alternatively, 'apopholasin');

(b) A sequence substantially homologous to or that hybridises to sequence (a) under stringent conditions; or

(c) A sequence substantially homologous to or that hybridises under stringent conditions to the sequence (a) or (b) but for the degeneracy of the genetic code; or

(d) An oligonucleotide specific for any of the sequences (a), (b) or (c).

The present invention will now be further described with reference to the accompanying Figures, in which:

Figure 1 shows three different cDNAs encoding apopholasin, referred to as clones 40, 3 and 5. Nucleotides in bold type show codons used for initiation and termination of translation;

Figure 2 shows the three sequences of Figure 1 aligned to demonstrate the sequence similarity. This figure was generated by Clustal. Positions which are indicated with a star are identical in all three clones. The codons for the initiation and termination of translation are highlighted in bold;

Figure 3 shows the oligonucleotides used for the complete sequencing of the positive clones. These were identified from the cDNA library; their positions in clone 40 are shown. Oligonucleotides are shown in bold type, portions of the flanking sequence of the Bluescript plasmid are shown in italic type;

Figure 4 describes the protein sequence described by the DNA sequence coding for apopholasin and shows, in Figure 4A, the complete sequence of the positive clone 40 identified from the *Pholas dactylus* light organ library. The first 20 amino acids at the N-terminus are a signal peptide, and this can either be retained or removed when generating the BIOP as described in this invention and, in Figure 4B, the cDNA coding for apopholasin with untranslated 5′ and 3′ ends. The untranslated regions are also shown;

Figure 5 describes the protein sequence for pholasin with (5B) and without (5A) the signal peptide;

Figure 6 shows the sequence for apopholasin genomic DNA. Two gDNA clones were indentifed but no introns were found; the Figure shows an alignment of the cDNA from cDNA clone 40 and the gDNA amplified by both *rTth* DNA polymerase XL and BioXAct polymerase. The sequences of the PCR product and the inserts in pGEM T were aligned with the sequence of the cDNA of clone 40 and were identical to this cDNA;

Figure 7 describes the oligonucleotides used for screening and expression. Degenerate oligonucleotides for library screening are shown in Figure 7A; non-degenerate ones in Figure 7B; and oligonucleotides used for protein expression are shown in figure 7C;

Figure 8 lists the main restriction sites in the DNA for engineering pholasin; and

Figure 9 is a schematic representation of Figure 8 mapped to the sequence of Figure 4A (translated region).

Accordingly, the present invention provides recombinant DNA encoding the apophotoprotein apopholasin and comprising the nucleotide sequence of the sequence disclosed in Figure 4B. Three different cDNAs coding for apopholasin have been isolated, having differing non-coding regions, respectively disclosed in Figure 1. The genomic DNA (gDNA), which contains no introns, has been shown (Figure 6) to comprise the same basic sequence as the cDNA.

Pholasin is a glycoprotein having 11.1 glusamine, 9.8 fructose, 7.1 mannose and 5.2 galactose residues. The cDNA for apopholasin has a molecular weight of 23,456 compared to 34,600 of the pholasin extracted from *Pholas*. The difference in the molecular weights of native versus recombinant apopholasin is due to the glycosylation of the native protein and luciferin. The isoelectric point of the translated protein calculated by the ISOELECTRIC command of the GCG programme is at 3.84. The native protein has a lower isoelectric point (<3.5), probably due to the presence of bound sulphate.

The three clones (Figure 2) isolated from the library encode a unique protein (Figures 4 and 5), which does not have the same amino acid sequence as any known protein in the SwissProt database. The present invention therefore not only provides cDNA and RNA coding for the protein, but also the recombinant protein *per se*, with or without glycosylation units. A comparison of segments of the pholasin protein sequence with the proteins in the SwissProt database identified several proteins with regions having a high sequence similarity to regions of the cloned protein. These included several proteins which interact with nucleotides (Table 1).

**Table 1** *A comparison of sections of the sequence of the cloned protein with sections of proteins which interact with nucleotides.*

| Protein | Homologous region<br>cloned protein<br>homology (+ denotes a conserved amino acid)<br>selected protein |
|---|---|
| tRNA-splicing endonuclease β subunit<br>*Saccharomyces cerevisiae*<br>EC 3.1.27.9 | SLYDEDNNGVMDEGKVIPSETIE<br>+L DEDNN + + G ++P E++E<br>NLRDEDNNLLDENGDLLPLESLE<br><br>LDQDVELDYTW<br>LD DV DYTW<br>LDHDVSKDYTW |
| ATP-AMP transphosphorylase<br>*Cyprinus carpio*<br>EC 2.7.4.3 | VMDEGKVIPSETIEDDIKDCGLLDQDVELDY<br>+M +G+++P +T+ D IKD + DV Y<br>IMQKGELVPLDTVLDMIKDAMIAKADVSKGY |
| DNA primase<br>*Synechocystis sp.*<br>EC 2.7.7. - | EEVQCAMNWTQANEYVFNVD<br>++VQ M ++Q+ + +FN D<br>DQVQSLMRFSQSKQIIFNFD |
| purine permease<br>*Emericella nidulans* | VQCAMNWTQANEYV<br>+ C+++WT+ N ++<br>IMCSVDWTRRNRFI |
| DNA repair protein complementing XP-A cells homologue<br>*Drosophila melanogaster* | PDTVDEAEDTPSET<br>PDT DE EDT + T<br>PDTYDEEEDTYTHT |
| ATP synthase β chain<br>*Peptococcus niger*<br>EC 3.6.1.34 | DTVDEAEDTPSET<br>D +DEA + PSET<br>DPIDEAGEVPSET |
| DNA polymerase α<br>*Homo sapiens* | DEDNNGVMDEGKVIPSETIEDDIKD<br>D+D G +++G+ I + +EDD D |

| | |
|---|---|
| EC 2.7.7.7 | DDDGIGYVEDGREIFDDDLEDDALD |

Similarity was found between the *Vargula* luciferase and *Renilla* LBP, but no other bioluminescent protein.

Sequence homology between the cloned protein and (a) *Vargula* luciferase (b) *Renilla* LBP. An area of high homology in all three proteins is in bold print.

(a) 148 206

GTIVVT**VRVSLYD**EDNNGVMDEGKVIPSETIEDDIKDCGLLD-QDVELDYTWTQNECDL

**V+VSL D** + + + T+ D I D + V++ + +

YWNTWD**VKVSLRD**VESYTEVEKVTIRKQSTVVDLIVDGKQVKVGGVDVSIPYSSENTSI

353 412

(b)

105 166

STMPGTYMLMDVCATRDADDKCIEGTIVVT**VRVSLYD**EDNNGVMDEGKVIPSETIEDDIKDC

+ TR + **VR+S+** + N+ K I

AIKIAKLSAEKAEETRGFLRVADQLGLAPG**VRISVEE**AAVNATDSLLKMKAEEKAMAVIQSL

41 104

Three potential glycosylation sites on the protein have the consensus triplet sequence Asn-Xaa-Ser/Thr (where Xaa can be any residue except proline). Thr 216 was identified as a potential site of O-linked glycosylation by a neural network which has been trained to identify this type of glycosylation. The amino acid sequence was also entered into a neural network which had been trained to identify eukaryotic signal peptides. This confirmed that the most likely cleavage site is between positions 20 and 21 (GSG-EE).

Many families of proteins contain a "signature" sequence of amino acids. The sequence of the clones did not contain any of these signatures present in the PROSITE database. The amino acids from 170 to 185 correspond to the calcium binding consensus sequence [DENQST]X[DENQST]X[DENQST]X[DENQST]X[DENQST]XX[DENQST]. Thirteen potential phosphorylation sites were discovered that matched the consensus sequences for either the kinase phosphorylation site [RK](2)-x-[ST], the protein kinase C phosphorylation site [ST]-x[RK] or the casein kinase II phosphorylation site [ST]-x(2)-[DE].

Three N-linked glycosylation sites were identified in the translated sequence of the clones A neural network has been trained to identify this type of glycosylation which identified Thr 216 as a potential site of O-linked glycosylation. At least one of these sites must be glycosylated in the native protein in order to account for the presence of the sugar residues. A putative signal peptide region preceded the N terminus of the secreted protein (determined by amino acid sequencing and was identified as a signal peptide by a neural network). To confirm this result the protein sequence was searched with PSORT for motifs which would locate the cloned protein in a cellular compartment. The protein sequence did not contain any transmembrane regions or N-myristoylation patterns which would indicate the presence of a lipid anchor. No targeting or retention sequences were found for the nucleus, mitochondria, endoplasmic reticulum or peroxisome.

The fact that the clones had some sequence similarity with proteins that interact with nucleotides may suggest that pholasin binds a co-factor as part of the chemiluminescent reaction. Beetle luciferases require ATP binding for chemiluminescent activity. There is no P-loop binding motif ((A,G)x4GK(S,T) or (A)x{4}GK(T)) in the amino acid sequence of these clones. However, not all ATP binding proteins contain this motif. Neither does the cloned protein contain the GXGXXG phosphate binding consensus sequence necessary for the binding of other co-factors such as nicotinamide adenine dinucleotide.

The amino acid and sugar components of pholasin are not able to emit light at the wavelength of the native protein (490nm). This indicates that there must be a chromophore bound to the protein. There are, however, proteins in which the chromophore is composed of modified amino acid residues within the polypeptide. The best characterised of these is the green fluorescent protein (GFP). This has a chromophore which is a ring formed by the autocatalytic cyclisation of the residues Ser-dehydroTyr-Gly. The serine may be mutated to a threonine, which increases the amplitude of the emission at 488nm. Pholasin had no similar amino acid sequence. Putative luciferin binding regions have been identified for two bioluminescent chemistries. Aequorin has a putative coelenterazine binding region, which is also present in two sections of the *Vargula hilgendorfii* luciferase. The sequence of the cloned protein has no homology to the putative luciferin binding site of aequorin, but the region of the *Vargula* luciferase from residue 353 to 411 has some similarity, as does the LBP of *Renilla reniformis*,

which also binds an imidazolopyrazine. This may indicate that the chemistry of pholasin bioluminescence involves an imidazolopyrazine luciferin. However, the region of homology is very small. The beetle luciferases contain an area of low sequence homology which may bind the benzothiazole luciferin. This low homology may account for the different colours of beetle bioluminescence. used a luciferin analogue (2-(4-benzoylphenyl) thiazole-4-carboxylic acid which photoinactivated the luciferase active site of the firefly *Photinus pyralis*. This photoinactivation was directly linked to the degradation of a small peptide sequence HHGF (residues 244-257). This is therefore postulated as a luciferin-binding site. The cloned protein does not have any sequence homology with these putative binding regions. Two strongly conserved regions of amino acids have also been found in both the luciferase and the luciferin binding protein of the dinoflagellate *Gonyaulax polyedra*. These regions were compared to the cloned protein, but no sequence similarity was found. No sequence identity could be established between the bacterial luciferases and the cloned protein.

Therefore, the present invention provides cloned apophotoprotein apopholasin (and the cDNA coding therefor), which has identical properties to native (but non-glycosylated) apopholasin with respect to molecular weight, amino acid composition, potential for glycosylation, its highly acidic pI and its cellular location. Hence, the present invention can further provide the corresponding BOIP or modified BOIP, according to standard methods.

The corresponding BOIP is preparable by bringing the apophotoprotein pholasin into association with the luciferin, also using standard methods. Although the luciferin is tightly bound in the native pholasin BOIP, it has been found that it may not be the case in the recombinant pholasin BOIP; indeed the luciferin may be weakly bound or merely present with the apoprotein. For example, a methanol, aqueous, acidic or other extract of *Pholas dactylus* (whole organism or light organ dissected from the animal) containing the 'luciferin', or the pure luciferin, is added to the solution, cell or organism. The luciferin associates with the apo-BOIP forming the photoprotein. The luciferin on the photoprotein then reacts with oxygen or one of its metabolites to produce light, in the presence or absence of the luciferase. The light emission may be detected, quantified, or imaged using a luminometer, photographic film or imaging camera, or by the naked eye. Alternatively, light emission may be generated spontaneously by intra- or extra-cellular metabolites

reacting with the apo-BOIP.

Although illustrated with respect to pholasin, the following may apply to any BOIP: the BOIP can be produced directly from native DNA, or from DNA engineered or amplified by the polymerase chain reaction. By this means, sites can be inserted within the protein by splitting the DNA into two or more pieces, or by adding DNA sequences to the 5′ or 3′ ends. For example, the DNA may be expressed in bacteria, yeast, an insect or human cell, or other suitable organism to produce protein which can be extracted and used.

In this instance, the protein produced from the cloned DNA reacts with oxygen or a metabolite of oxygen, such as the superoxide anion ($O_2^-$), hydrogen peroxide ($H_2O_2$), hydroxy radical ($OH^-$), an oxyhalide anion ($OCl^-$, $OBr^-$, $OI^-$, $OSCN^-$), nitric oxide (NO), an organic hydroperoxide or a radical $ROO^-$. The change in light emission enables the oxygen or metabolite(s) to be detected and quantified in live cells, organelles, or on the outer or inner surface of the plasma membrane, or within an organ of a live organism without the need to break them open or the need to separate bound and free fractions. This also enables an enzyme producing oxygen or one of its metabolites, such as chlorophyll, or enzymes such as oxidases and oxygenases which react directly with oxygen or one of its metabolites to attach oxygen to the substrate to be detected and quantified in live cells, organs and whole organisms, or extracts from any one of these.

Also the BOIP can be made *in vitro* by transcription/translation in a cell lysate such as rabbit reticulocyte lysate or wheat germ extract containing RNA polymerase. The DNA for the BOIP is first engineered to contain an RNA polymerase promoter, such as T7, SP6; bacterial promoter(s), such as lac, ara or trp; or mammalian promoter(s), such as actin, myosin, myelin proteins, TK, MRT-V, SV40, CMV, RSV, metallothionine, antibody, G6P dehydrogenase, and can be amplified *in vitro* using the polymerase chain reaction. A poly-A tail may be added at the 3′ end and a tissue specific promoter or enhancer sequence added to the 5′ or 3′ end of the DNA coding for the BOIP or modified BOIP, enabling it to be expressed specifically in a target cell, such as a myocardial cell or a cancer cell. The expression of the BOIP in the target cell is detected and quantified by light intensity, colour or polarisation, as previously mentioned.

The BOIP, or its DNA or RNA, may be incorporated into a live bacteria or eukaryotic cell using phage, virus, plasmid, calcium phosphate transfection, electroporation, liposome fusion, membrane pore forming proteins, micro-injection or DNA gun. Once inside cells or an appropriate extracellular environment, cell activation or injury will initiate or change the light emission from the BOIP. For example, expression in live organisms by micro-injection of protein, RNA or DNA, or by transgenic manipulation to produce a cell, such as a bacterial, microbial, animal or plant cell, *eg* a white blood cell, a heart cell, or a yeast, protozoan, fruit fly (*Drosophila*), nematode worm, polychaete worm, fish, human, mouse, rat, sheep, pig, horse or plant, which can generate its own light.

The BOIP can also be incorporated into a defined part of a live cell by chemical means or by genetically engineering the BOIP to contain a signal peptide, which locates the BOIP to the inner or outer surface of the plasma membrane or within a particular organelle, such as peroxisome, mitochondrion, chloroplast, tonoplast, endoplasmic, reticulum, Golgi apparatus, endosome, lysosome, secretory vesicle, nucleus, nucleolus, nuclear membrane, plasma membrane, proteosome, or gap junction, or structure such as membrane receptor ion channel microtubule, cytoskeleton, nuclear skeleton, nuclear receptor, mitotic spindle or microfilaments. The signal peptide, added either chemically or genetically, will normally target the normal or modified BOIP to a particular intra- or extra-cellular site. For example, the sequence MLSRLSLRLLSRYLL or part of cytochrome oxidase on the N-terminus will target the BOIP to the mitochondrion; KKSALLALMYVCPGKADKE or MLLPVPLLLGLLGLAA at the N-terminus will target the BOIP to the endoplasmic reticulum, a KDEL or HDEL or KEEL sequence at the C-terminus retaining it there. SKL at C-terminus targets BOIP to the peroxisome; PKKKRKV or an extension of this SV40 large T-antigen signal will target it to the nucleus; and a palmitoylation and/or a myristoylation signal will target it to the plasma membrane. By coupling the BOIP to another protein that targets itself to a particular site, the BOIP can also be targeted there. For example, coupling the nuclear proteins nucleoplasmin or lamin B receptor to BOIP targets it to the nucleus; cytochrome oxidase at the N-terminus targets BOIP to the mitochondria; chlorophyll at the N-terminus targets BOIP to the chloroplast; a connexin at the N-terminus targets BOIP to the gap junction or plasma membrane; and SNAP 25 to the plasma membrane.

Other modifications to the apoprotein, BOIP, or nucleotides coding therefor include, but are not limited to:

The apoprotein, such as apopholasin, may also be glycosylated, and used to detect and quantify secretion or movement of proteins through the secretory pathway.

Nucleic acid coding for the BOIP when expressed inside a live cell may not only be modified but also regulated in this cell by gene expression, such as by promoters, enhancers or oncogenes. For example, the apoprotein, such as apopholasin, may be coupled to a gene regulator protein, such as a transcription factor, by genetic or chemical manipulation, such that the movement through a cell or of the regulator protein or its activity, can be detected or quantified.

The BOIP or apoprotein, or its DNA may be linked to another protein or DNA used in therapy, such that the other protein or DNA can be detected in live cells or in a whole organism, *eg* a human.

The apoprotein, such as apopholasin, can also engineered genetically or chemically to contain a site or sites which can be covalently modified by enzymes such as phosphorylation (including ser/thr, his and tyr kinases and phosphatases), tranglutamination, proteolysis, ADP ribosylation, gly-or glu-cosylation, halogenation, oxidation, methylation, palmitoylation, myristylation and farnesylation.

The apoprotein, such as apopholasin, can be engineered genetically or chemically to contain an antigen or intracellular signal binding site, such as $Ca^{2+}$, cyclic AMP, cyclic GMP, cyclic CMP, $IP_3$, $IP_4$, diacyl glycerol, ATP, ADP, AMP, GTP, or any oxy- or deoxy-ribonucleoside or nucleotide, a substrate, a drug, a nucleic and/or a gene regulator protein.

The BOIP may also be converted to a rainbow protein by engineering a particular site such as described hereinabove into the BOIP, at the N- or C-terminus, or between a chimera of the BOIP and an energy transfer acceptor, such as GFP (wild type or any of the mutant

GFPs). This is known as chemiluminescence, bioluminescence or fluorescence resonance transfer (CRET, BRET or FRET, respectively). Conversion of the BOIP to a 'rainbow protein' may be effected by reaction with a cellular substance, modification genetically or chemically, or by linking the BOIP to a fluor, such as the green fluorescent protein or the red fluorescent protein in the deep sea fish *Malacosteus*. The result is a BOIP which changes its colour and/or intensity and/or polarisation of emission. The change in colour occurs by energy transfer, *eg* resonance transfer (CRET or FRET) or electron transfer.

The initial (unmodified) BOIP may be the apophotoprotein, its DNA or RNA, from the bivalve mollusc *Pholas dactylus*, or it may be another BOIP, such as one from the mollusc *Rocellaria grandis* or the squid *Ommastraphes*, or earthworm luciferase, which produce light with oxygen metabolites in a way very similar to *Pholas dactylus*.

The BOIP, apo-BOIP, or nucleic acid coding for it, whether modified or not, may therefore be used in a range of biology and investigations such as:

(a) Detection, location and measurement of signals in substrates, such as live cells, organs or organisms, or in extracellular fluids;

(b) Detection, location and measurement of oxygen and its metabolites in substrates, such as live cells, organs or organisms, or in extracellular fluids, water (sea and fresh), soil or the atmosphere;

(c) Detection and location of normal cells such as microbes (protozoa, yeast, fungi, moulds, bacteria, viruses);

(d) Detection and location of abnormal cells, such as cancer cells, hyperactive cells in rheumatoid arthritis and other inflammatory diseases, cells infected with a pathogen, such as a virus or other infectious agents, cells damaged by physical, chemical or biological attack, cells damaged by perfusion or reperfusion injury or cells damaged by oxygen or one of its metabolites;

(e) Measurement and location of enzymes, particularly those producing oxygen or its metabolites, and other tumour reactions in cells or biological fluids;

(f) DNA and RNA binding assays;

(g) Immunoassay and other protein binding assays;

(h) In genetic engineering, in the development of transgenic animals and plants, and

microbes; in horticulture; agriculture; medicine and veterinary medicine; and/or

(i) in genetic entertainment by incorporation into light sticks, greeting cards or toys to produce light of various colour, intensities, oscillations, flashes and glows; or in comestibles, such as food, drinks, including beers, wines, spirits, colas and other soft drinks.

Accordingly, the present invention further provides an apoprotein, such as pholasin apoprotein (or apopholasin) in both unglycosylated and glycosylated forms, and a BOIP thereof, such as pholasin, either alone (but excluding native proteins *per se* that have already been isolated, such as native pholasin *per se*) or in association with one or more of: a targeting or signal peptide; a glycosylate; a site capable of modification by an enzyme; an antigen or intracellular signal binding site; a promoter, an enhancer or an oncogene or a pharmacologically active substance; or the like. The present invention further provides a recombinant construct comprising a nucleic acid sequence encoding for any of these proteins; a vector containing a nucleic acid sequence encoding for any of these proteins; a host transformed by such vector; a live cell, such as bacterial, insect, eukaryotic, prokaryotic, archae or plant cells containing or expressing any of these proteins; and a rainbow protein, as described herein, together with a nucleic acid sequence encoding therefor.

The present invention will now be illustrated with reference to the following non-limiting examples, in which the methodology referred to is known to those skilled in the art and/or may be carried out by analogy with reference to the protocols disclosed in the following references, the contents of which are herein incorporated by reference in their entirety:

**BOOKS**

1. Campbell,AK. (1988). Chemiluminescence: principles and applications in biology and medicine, pp608. Horwood/VCH, Chichester and Weinheim.

2. Campbell,AK. (1994). Rubicon: the fifth dimension of biology. pp 304. Duckworth, London.

**PAPERS**

1. Campbell,AK Patel,A. (1983) Biochem J. 216:185-194. A homogeneous immunoassay for cyclic nucleotides based on chemiluminescence energy transfer.

2. Roberts,PA Knight,J Campbell,AK. (1987) Anal. Biochem. 160:139-148. Pholasin - a bioluminescent indicator for detecting activation of single neutrophils.

3. Mueller,T Davies,EV Campbell,AK. (1989) J. Biolum. Chemilum. 3:105-113. Pholasin Chemiluminescence Detects Mostly Superoxide Anion Released from Activated Human Neutrophils.

4. Mueller,T Campbell,AK. (1989) J. Biolum. Chemilum. 5:25-30. The Chromophore of Pholasin: A Highly Luinescent Protein.

5. Sala-Newby,G and Campbell,AK (1991) Biochem.J. 279:727-732. Engineering a bioluminescent indicator for cyclic AMP dependent protein kinase .

6. Campbell, AK, Trewavas, AJ and Knight, MR (1996). Calcium imaging shows differential sensitivity to cooling and communication in luminous transgenic plants. Cell Calcium 19: 211-218.

7. Kendall, JM, Sala-Newby, G Badminton, M, Campbell AK and Rembold, CR (1996). Free $Ca^{2+}$ in the endoplasmic reticulum of living cells measured using aequorin as a pseudo luciferase. Biochem.J 318:383-387

8. Badminton, MN, Campbell, AK and Rembold, CR (1996). J.Biol. Chem. 271:31210-31214. Differential regulation of nuclear and cytosolic $Ca^{2+}$ in HeLa cells.

9. Sala-Newby, GB, Taylor, KT, Badminton, MN, Rembold, CR and Campbell, AK (1998). Imaging bioluminescent indicators shows $Ca^{2+}$ and ATP permeability thresholds in, live cells attacked by complement. Immunology. 93:4:601-609

Sala-Newby, GB, Kendall, JM, Jones, H, Taylor, KM, Badminton, MN, Llewellyn, DH and Campbell, AK (1999). Targeting bioluminescent proteins to defined compartments fo living cells. Methods in Enzymology Bioluminescence and Chemiluminescence. ed Ziegler, M and Baldwin, T.

**EXAMPLE 1: Production of a BOIP in bacteria**

c or genomic DNA coding for apopholasin, with or without the cDNA coding for the signal peptide, is amplified by PCR with restriction sites such as BamHI at each end. The cDNA is run on an agarose gel and the full length DNA eluted and purified. The DNA is then cut with BamHI to generate sticky ends and ligated into an expression plasmid such as pET3a, which has been cut with BamHI also. After ligation the sealed plasmid is transformed into a standard *E.coli* K12 strain such as JM109, a colony picked off for a large plasmid preparation. After checking that the plasmid contains the correct sequence for apopholasin and is in the correct orientation the plasmid is then used to transform a standard expression strain of E.coli such as BL21(DE3) or other expression strain. A colony is picked off the agar plate and grown up for 2h in standard LB broth. IPTG is added as inducer for a further 2h. Apopholasin can then be extracted by breaking the bacteria by lysozyme digestion or sonication in a standard salt medium such as 50mM HEPES pH 7 +/- lmM ascorbate. Since the apopholasin is unglycosylated it tends to aggregate and form inclusion bodies. These can be broken using 8M urea or guanidinium chloride and these then dialysed off. If the pH of PAGE gels is alkaline this also tends to allow aggregation of both the unglycosylated and glycosylated apo- and full pholasin. A signal peptide such β-lactamase signal will target the BOIP to the periplasmic space, resulting in the ability to secrete the expressed protein from the external fluid of the cells.

**EXAMPLE 2: Production of a BOIP in insect cells**

c or genomic DNA coding for apopholasin is inserted into a plasmid suitable for conversion into baculovirus when transfected into insect cells. Since pholasin is secreted by *Pholas* itself there is a signal peptide at the N-terminus. Removal of this by PCR will allow cytosolic expression in insect cells. If the signal peptide is left on or changed for honey bee mellitin signal peptide, the apopholasin is secreted into the external medium. The virus containing the DNA for apopholasin is then purified and stored until required. An aliquot is then added to fresh insect cells and these incubated for 3-7 days. The apopholasin is then isolated from the supernatant if a signal peptide is used, or from the cells is not. The apopholasin can then be purified by ammonium sulphate precipitation, gel filtration and DEAE chromatography. The state of glycosylation can be assessed by running the protein on PAGE when the molecular weight is 34Kda. Removal of the

glycosylation by enzymes returns the protein to the size of apopholasin 23.5Kda. It can be stored frozen or freeze dried, and activated to form pholasin by addition of luciferin as described in Example 3.

Since the apopholasin tends to aggregate in the insect supernatant it is important to get the protein into non-aggregating buffer, e.g. 50mM HEPES pH 6, 1-10mM ascorbate, as soon as possible.

Formation of pholasin can then be achieved as described in Example 3.

**EXAMPLE 3: Generating pholasin and light emission**

To generate light the apopholasin must first be converted into pholasin with the luciferin. The luciferin can be extracted from native pholasin by mild acid, or by methanol, mild acid or alkaline treatment of light organs isolated from *Pholas dactylus* or the whole organism. After homogenisation the extract is centrifuged or filtered to remove particulate material. Further purification can be achieved by tlc of hplc. The luciferin is best stored dry, but can be stored at -70°C. The intactness and concentration can be estimated by measuring the absorbance or fluorescence. The details are as follows:

(a) *Isolation of the luciferin*

Four protocols (1-4) have been developed to extract and isolate the luciferin responsible for light emission in pholasin. The luciferin is a small organic moiety tightly bound to apopholasin when pholasin is isolated from *Pholas dactylus*, but also can be found not bound to apopholasin. Thus the extraction procedure isolates either form of luciferin.

1. The organism *Pholas dactylus* or its light organs are homogenised in 50mM sodium phosphate pH 6.0 on ice. The pholasin is precipitated with saturated ammonium sulphate (4°C stirred), and then removed by centrifugation at *ca* 15,000g for 30min in the cold. The supernatant is then passed down a SEP-PAK silica column, which binds the luciferin. The column is washed with 5ml ethyl acetate and then 5ml of methanol. The active fractions containing the luciferin are assayed either by reactivation of the apopholasin or by

chemiluminescence in DMSO, DMF, or NaOCl. The luciferin is concentrated and can be purified further on tlc or hplc with a standard solvent. It is dried and stored at -70°C.

2. The organism *Pholas dactylus* or its light organs are homogenised in cold acetone on ice, filtered through a Buchner funnel, and extracted with methanol:acetone (1:1), the residual powder being extracted 3 times with methanol and extracts combined. These are then concentrated in a Rotavaporator and left to stand for 1h on ice to allow further precipitation. The suspension is then refiltered and concentrated. The solution containing the luciferin is then passed down a SEP-PAK silica column which binds the luciferin. The column is washed with 5ml ethyl acetate and then 5ml of methanol. The active fractions containing the luciferin are assayed either by reactivation of the apopholasin or by chemiluminescence in DMSO, DMF, or NaOCl. The luciferin is concentrated and can be purified further on tlc or hplc with a standard solvent. It can be dried and stored at -70°C.

3. The organism *Pholas dactylus* or its light organs are homogenised in cold acetone on ice, and filtered through a Buchner funnel to give an acetone powder. This is then extracted with methanol:acetone (1:1), twice for 10min and then 3 times with methanol. The extracts are combined and concentrated in a Rotavaporator. They are left to stand for 1h on ice to allow further precipitation, refiltered and concentrated. The residual powder is resuspended in 50mM sodium phosphate pH 6.0, 10mM ascorbate, and ultrafiltered with a 10kD Amicon membrane at 4 C for pholasin. The solution containing the luciferin is then passed down a SEP-PAK silica column which binds the luciferin. The column is washed with 5ml ethyl acetate and then 5ml of methanol. The active fractions containing the luciferin assayed either by reactivation of the apopholasin or by chemiluminescence in DMSO, DMF, or NaOCl. The luciferin is concentrated and can be purified further on tlc or hplc with a standard solvent. It is dried and stored at -70°C.

4. The organism *Pholas dactylus* or its light organs are homogenised in 50mM HEPES buffer, with methanol and 100mM HCl on ice, and incubated for 2h on ice. After centrifugation at *ca* 15,000g for 30min in the cold, the

supernatant is then passed down a SEP-PAK silica column which binds the luciferin. The column is washed with 5ml ethyl acetate and then 5ml of methanol. The active fractions containing the luciferin are assayed either by reactivation of the apopholasin or by chemiluminescence in DMSO, DMF, or NaOCl. The luciferin is concentrated and can be purified further on tlc with a standard solvent. It is dried and stored at -70°C.

Method 4 normally generates most luciferin. The luciferin is characterised by its absorbance and fluorescence spectrum, and by its chemiluminescence with DMSO, NaOCl and apopholasin.

(b) *Generation of pholasin from apopholasin and the luciferin*

A small sample of the luciferin (1-10µl) is added to apopholasin in an appropriate buffer (50mM HEPES pH 6-7.5, +/- 1-10mM ascorbate, or 500mM NaCl, 10mM TES, 1mM EDTA, 1mM mercaptoethanol pH 6-7.5). The mixture is incubated at room temperature for up to 24h, and the pholasin assayed by adding an oxygen metabolite, e.g. NaOCl, or luciferase to a sample. When apopholasin has been expressed in cells, the luciferin is added externally, microinjected into individual cells or added via liposomes to get the luciferin into the cell.

Light is detected and quantified in a standard luminometer, imaging camera (intensified or CCD), or by a silicon chip.

**EXAMPLE 4: Production of a BOIP in vitro**

c or genomic DNA coding for apopholasin, with or without the signal peptide, is amplified by PCR with the 5' primer containing the DNA coding for T7 RNA polymerase. The DNA product is purified and precipitated. After dissolving in 10mM tris/1mMEDTA pH7, the DNA is added to a standard *in vitro* transcription/translation system such as rabbit reticulocyte lysate or wheat germ agglutinin and incubated at 30°C for 30-60min. The apopholasin can then be purified and activated to form pholasin as described in Example 3.

**EXAMPLE 5: Targeting a BOIP in vitro**

The BOIP can also be incorporated into a defined part of a live cell by chemical means or by genetically engineering the BOIP to contain a signal peptide which locates the BOIP to the inner or outer surface of the plasma membrane or within a particular organelle such as

peroxisome, mitochondrion, chloroplast, tonoplast, endoplasmic reticulum, Golgi, endosome, lysosome, secretory vesicle, nucleus, nucleolus, proteosome, or gap junction, or structure such as microtubule, cytoskeleton, nuclear skeleton, nuclear receptor, or mitotic spindle. The signal peptide, added either chemically or genetically, will normally target the normal or altered BOIP to a particular intra- or extra-cellular site for example, the sequence MLSRLSLRLLSRYLL or part of cytochrome oxidase on the N-terminus will target the BOIP to the mitochondrion; KKSALLALMYVCPGKADKE or MLLPVPLLLGLLGLAA or the ER protein calreticulin at the N-terminus will target the BOIP to the endoplasmic reticulum, a KDEL or HDEL sequence at the C-terminus retaining it there. SKL at C-terminus targets BOIP to the peroxisome, PKKKRKV or an extension of this SV40 large T-antigen signal will target it to the nucleus, and a palmitoylation and/or a myristoylation signal (MGCVCSSNPD = the LCK N-terminal acylation motif from tyrosine kinase) will target it to the plasma membrane. By coupling the BOIP to another protein which targets itself to a particular site then the BOIP is also targeted here. For example, coupling the nuclear proteins nucleoplasmin or lamin B receptor to BOIP targets it to the nucleus; cytochrome oxidase at the N-terminus targets BOIP to the mitochondria; chlorophyll at the N-terminus targets BOIP to the chloroplast; and a connexin at the N-terminus targets BOIP to the gap junction or plasma membrane, SNAP 25 to the plasma membrane.

In order to target pholasin to defined sites in living cells, the DNA coding for these targeting sequences are added by using PCR. For cytosolic apopholasin the native signal peptide is removed and also the BOIP can be linked to larger proteins at the N- or C-terminus such as firefly luciferase or aequorin to prevent it getting into the nucleus. This also enables ATP and oxygen metabolites, or $Ca^{2+}$ and oxygen metabolites to be measured simultaneously in the same cells by intensity, colour or polarisation of the different bioluminescent indicators. A multiple bioluminescent indicator can also be engineered by PCR, or by using restriction enzyme sites, from the DNA coding for 3 or more bioluminescent proteins. A simple screen of the transformed bacteria enables the multiple rainbow protein to be isolated with 2-3 colours or more.

The DNA is then added to an *in vitro* transcription/translation system as described in Example 4 in the presence of the organelle to be targeted (e.g. microsomes for the endoplasmic reticulum, which glycosylate apopholasin).

The new DNA can also be inserted into a plasmid by standard techniques, and transformed into bacteria or transfected or injected into eukaryotic cells such as HeLa or COS.

Addition of the luciferin as described in Example 3 allows formation of pholasin which can then be detected by light emission. Changes in oxygen metabolite production are then be detected by a luminometer or imaging camera when the cells are exposed to external oxygen metabolites, a change in oxygen concentration, addition of stimuli e.g. TNF, EGF, hormones or drugs, or attack by pathogens such as bacteria, viruses, complement, antibodies, toxins, and cells of the immune system.

**EXAMPLE 6: Engineering a covalent modification site into a BOIP**

(a) The site coding a protein kinase A (RRAS or kemptide), protein kinase C (MARCKS), MAP kinase, ERK, the ER - nuclear signalling kinase IRE1P or a phosphatase is added to the N- or C-terminus or inserted at various sites within the apopholasin by PCR and expressed as described in Examples 1-5. Pholasin is then generated by addition of the luciferin as described in Example 3.

Addition of the catalytic subunit for protein kinase A, or activation via cyclic AMP inside cells, leads to phosphorylation or dephosphorylation of the modified pholasin and change in light emission (intensity, colour or polarisation).

A preliminary screen is necessary to select the appropriate proteins and to discard any which have lost all activity.

(b) The site coding a protease (thrombin, enterokinase, HIV protease, caspase) is added to the N- or C-terminus of the apopholasin by PCR or inserted at various sites within the protein, and expressed as described in Examples 1-5. Pholasin is then generated by addition of the luciferin as described in Example 3.

## EXAMPLE 7: Engineering a BOIP into a "Rainbow Protein"

cDNA coding for apopholasin is linked to another protein by using the cDNA coding for that protein. For example, wild type GFP, the S65T mutant of the green fluorescent protein, YGFP, or EGFP are linked to apopholasin by PCR at the N- or C-terminus, or by splitting one or both proteins using multi-step PCR. In between there is a 'reactive' peptide with a protease site (α thrombin or enterokinase) and a binding site for $IP_3$, or the 15 amino acid sequence form $IP_3$ kinase (an $IP_4$ binding site). At the C-terminus of the GFP, a peptide containing 6 lysine residues may also be added via PCR. The protein is expressed and fluorescein covalently linked to these lysines by addition of fluorescein isothiocyanate. Addition of the luciferin forms pholasin as described in Example 3. The change in colour occurs by chemiluminescence resonance energy transfer. Without fluorescein the rainbow protein emits blue-green light (508nm), which changes to blue (490nm) when the reactive substance binds to the reactive peptide, or when either thrombin or enterokinase is added. When the 6 amino acid linker is used the colour starts as green (530nm), and will then change from green, to blue-green and then blue as the particular reactive sequence binds their respective analytes. Use of rhodamine instead of fluorescein generates a rainbow protein which changes from red to green to blue.

A preliminary screen is necessary to select the appropriate rainbow proteins and to discard any which have lost all activity.

The other protein linked to apopholasin may be, for example, any one of the following linked chemically or genetically:

1. Firefly or any benzothiazole luciferase to the N or C terminus gives two colours for ATP and oxygen metabolites.
2. Any imidazolopyrazine luciferase, including coelenterazine systems - decapod shrimp, fish, sqiud, *Renilla*, anthzoan, Chaetognate, radiolarian, or copepod and *Vargula* systems - ostracod, *Porichthys* and similar fish, cypridinids and *Vargula*.
3. Any tetrapyrrole luciferase such as dinoflagellate, euphausiid or stomiatoid fish.

4. Bacterial luciferase and other aldehyde or flavin luciferases, including polychate worm.

5. Any GFP, including wild type, S65T, enhanced GFP, blue GFP, yellow GFP, *Renilla* GFP, *Ptilocarpus* GFP, and Pennatula GFP, any anthozan GFP, or any coelenterate GFP.

6. The red fluorescent proteins from stomiatoid fish - *Malactosteus*, *Aristostomias*, *Photostomias*.

7. The phycobiliproteins - phycoerythrin and phycocyanobilin.

8. The blue fluorescent lumazine protein in the bacterium *Photobacterium*.

9. The yellow flavin fluorescent protein in Y *Vibrio*.

10. Any lysine or argininine or other amino acid side chain where a fluor can be added covalently. In which the case the rainbow protein amy emiot more than two colours. For example, rhodamine on a pholasin-linker-GFP chimera will turn from red to green to blue.

A preliminary screen may be necessary to select chimeras which have not lost all bioluminescent activity.

The 'reactive' peptide may be a binding site for any analyte, protein or DNA, metabolite, substrate vitamin, an enzyme such a protein kinase C or phosphatase, ion channel, ion pump, antigen, antibody, nucleotide or nucleoside such as ATP, GTP, ADP, AMP, adenosine, cAMP, cGMP, cCCP or their deoxy equivalents, and inositol phosphates such as $IP_3$ or $IP_4$, a lipid such as diacyl glycerol, phosphatidyl inositol bisphosphate, phosphate, a cation such as $Ca^{2+}$, $K^+$ or $Na^+$, $Cu^{2+}$ or $Zn^{2+}$, or anion such as $Cl^-$, sulphate, or gas such as NO, $O_2$ or $H_2$, or a protein binding site such as calmodulin, kinesin, dynein, tubulin, or myosin.

When pholasin is triggered by oxygen metabolites, the *Pholas* luciferase or peroxidase, energy transfer occurs from pholasin oxyluciferin through GFP to fluorescein resulting in a yellow emission. Addition of thrombin for 3h cleaves the GFP-fluorescein from the pholasin and the light emission returns to the blue of native pholasin. Addition of $IP_3$ to the full chimera alters the efficiency of energy transfer. As a result there is a change in the ratio of light emitted in the yellow to blue. This ratio is directly related and can be plotted

against the concentration or amount of analyte. The light is detected in a dual wavelength luminometer or ratiometric imaging camera and the ratio of blue to green light measured.

Alternatively any fluors can be used, and any binding sites with the right characteristics as shown in these examples will work provided a simple screen is used to select the right chimeras.

**EXAMPLE 8: Engineering a BOIP into a "Rainbow Protein" for two analytes together**

Apopholasin is linked to firefly luciferase by using cDNAs and PCR, followed by expression in insect cells as described in Example 2. Addition of the luciferin as described in Example 3 generates the pholasin. In the presence of firefly luciferin (1mM), ATP and oxygen metabolites, this chimera emits blue and yellow simultaneously which can be independently measured by using a dual wavelength luminometer or imaging camera.

**EXAMPLE 9: Expression of BOIPs in mammalian cells**

Apopholasin, c or genomic, in an expression plasmid with the CMV promoter, is transfected into HeLa cells. After incubation for 3 days to allow expression of the apopholasin, the luciferin is added to form pholasin. Expression is checked using a polyclonal antibody to pholasin raised in rabbits. Addition of oxygen metabolites outside the cell allows the permeability of the plasma membrane to oxygen metabolites to be assessed. As the oxygen metabolites permeate into the cytosol, the light emission increases.

**EXAMPLE 10: Expression of BOIPs in plants**

c or genomic DNA coding for apopholasin is inserted into a plasmid with the cauliflower mosaic virus promoter and transformed into *Agrobacterium tumificans*. These are then added to a tobacco leaf, seedlings generated, and those expressing apopholasin selected. The plants are grown to seed, and seedlings grown from this seed. Addition of luciferin forms the pholasin as described in Example 3. Stressing the plant, e.g. with wind, touch, cold, or peroxide, or during growth and development or by a hormone, generates light, showing the formation of oxygen metabolites within the live plant. A cell-specific promoter engineered on to the apopholasin cDNA before making the transgenic plant enables oxygen metabolites to be detected in specific cells within the whole, living plant.

**EXAMPLE 11: Detection of oxidative damage in vitro**

Addition of pholasin to serum or plasma from a rat, mouse or human enables oxygen metabolites to be detected and measured on addition of a drug or other substance of interest.

**EXAMPLE 12: Detection of ROMs in a heart cells**

Reperfusion has been proposed to lead to oxygen metabolite damage in cardiac myocytes. Pholasin allows this to be tested for the first time. Plasmid containing apopholasin cDNA and the CMV promoter is transfected into isolated cardiac myocytes in culture. Expression occurs within 1-3 days, and pholasin is formed by addition of the luciferin as described in Example 3. Subjecting the cells to hypoxia followed by readmission of normal oxygen leads to light emission, showing that oxygen metabolites have been generated inside the cells. By using an imaging camera, the digital or analogue nature of this can be assessed as the number of cells emitting light can be visualised and counted.

**EXAMPLE 13: Detection of ROMs in the nucleus and endoplasmic reticulum (ER)**

Plasmid-containing apopholasin cDNA with either nucleoplasmin DNA or calreticulin DNA (with or without KDEL on the C-terminus) linked to the pholasin DNA, to target the apopholasin to the nucleus or ER respectively, and the CMV promoter for expression, is transfected into HeLa cells in culture. Expression occurs within 1-3 days, and pholasin is formed by addition of the luciferin as described in Example 3. Addition of oxygen metabolites outside the cells, or hypoxic/oxygen shock generates light measured in a luminometer, showing how fast oxygen metabolites penetrate into these organelles. By imaging with a photon counting imaging camera, the number of cells permeable to oxygen metabolites can be counted. Location of the pholasin can be assessed by imaging live cells, or by using immunofluorescence with the pholasin antibody on partially-fixed cells or GFP-pholasin in live cells. Using a rainbow protein, two or more analytes can be detected together.

**EXAMPLE 14: Use of pholasin as a protein label**

Pholasin can be used as a label in homogeneous or heterogeneous immunoassay. Apopholasin is first covalently linked to an antibody to HIV, and pholasin formed by addition of luciferin as described in Example 3. The antibody is then used in a standard chemiluminometric immunoassay format. Addition of HIV antigen leads to an increase in antibody binding and an increase in light emission dependent on the amount of HIV added. The amount of HIV in a blood sample can be assessed by relating the pholasin light emission in the sample to the standard curve.

**EXAMPLE 15: Pholasin as a DNA label**

Apopholasin is covalently linked to an oligonucleotide probe for detecting the presence of the cystic fibrosis gene. Addition of the probe to DNA in a standard Southern blot allows the probe to bind when the gene is present. Addition of luciferin as described in Example 3 allows the pholasin to form. Addition of hypochlorite (10mM) in barbitone buffer pH 9 causes the pholasin to flash and the gene can be visualised by the photon counting imaging camera.

**EXAMPLE 16: Pholasin in a two hybrid system**

Protein-protein interaction can be detected by engineering apopholasin to one half of a two hybrid system and GFP to the other. Binding will allow the yeast to grow.

**EXAMPLE 17: Pholasin in genetic entertainment**

Pholasin is able to chemiluminesce at a wide range of pH (3-10), including acid pH such as 3-4. Thus it can be added to drinks such as beer, cola, soft drinks, and spirits to make them glow. It can also make food glow by adding to them to the ingredients of cakes, icing, popcorn; by painting the pholasin or apopholasin on to the food, or by making it genetically in the source of the food. It can be used in a wide range of toys and other entertaining devices including squirt guns, greeting cards, pens.

The rainbow proteins can also be used as an alternative to pholasin alone, resulting in a rainbow of colours and colour changes.

**EXAMPLE 18: Pholasin in trangenic animals**

Transgenic animals such as nematodes, mice or plants can be generated from apopholasin cDNA by standard techniques. Injecting the luciferin or incubating whole plant in it forms the active pholasin. Oxygen or its metabolites can then be detected, measured and imaged, in an intact organ, or from the whole organism. It can also be used in humans, in DNA therapy or diagnosis.

**EXAMPLE 19: Apoprotein from the luminous squid Ommastrophes**

The use of apoprotein from the luminous squid *Ommastrophes* is substituted for apopholasin, and the methods of Examples 1 to 18, above, are carried out.

**EXAMPLE 20: Apoprotein from the mollusc Rocellaria**

The use of the apoprotein from the mollusc *Rocellaria* is substituted for apopholasin, and the methods of Examples 1 to 18, above, are carried out.

**EXAMPLE 21: Earthw rm luciferase**

The use of earthworm luciferase as a BOIP is substituted for apopholasin, and the methods of Examples 1 to 18, above, are carried out.

**EXAMPLE 22**

Genomic DNA from *Pholas*, *Rocellaria*, *Ommastrophes*, or earthworm is substituted for the recombinant protein in Examples 1 to 18, above, the methods of which are carried out in an analogous manner.

## CLAIMS

1. An isolated, purified or recombinant nucleic acid sequence comprising:

(a) a sequence that encodes the apophotoprotein of pholasin (alternatively, 'apopholasin');

(b) a sequence substantially homologous to or that hybridises to sequence (a) under stringent conditions; or

(c) a sequence substantially homologous to or that hybridises under stringent conditions to the sequence (a) or (b) but for the degeneracy of the genetic code; or

(d) an oligonucleotide specific for any of the sequences (a), (b) or (c).

2. A sequence according to claim 1, wherein the sequence that encodes for apopholasin is as shown in Figure 4B.

3. A sequence according to claim 1, wherein the sequence that encodes for apopholasin is as shown in any one of Figures1, 2, 3, 4A, 6 or 9.

4. A sequence according to any preceding claim, wherein the apopholasin is non-glycosylated.

5. A sequence according to any preceding claim, wherein the apopholasin is glycosylated.

6. An isolated, purified or recombinant construct incorporating a sequence encoding apopholasin protein according to any preceding claim.

7. An isolated, purified or recombinant construct incorporating a sequence encoding an apophotoprotein whose expression in a substrate, in association with a luciferin therefor, signals the presence of oxygen or an oxygen metabolite in the substrate.

8. An isolated, purified or recombinant construct incorporating a sequence encoding an apophotoprotein whose expression in a substrate, in association with a luciferin therefor, signals the presence of oxygen or an oxygen metabolite in the absence of a corresponding luciferase in the substrate.

9. A recombinant construct according to any one of claims 1 to 8, wherein the nucleic acid sequence is linked operably with nucleotides enabling expression and secretion of the apopholasin in a cellular host.

10. DNA or RNA according to any of claims 1 to 9.

11. An isolated, purified or recombinant polypeptide comprising apophotoprotein of pholasin (apopholasin) or a mutant or variant thereof having substantially the same

activity as apopholasin.

12. An isolated, purified or recombinant polypeptide according to claim 11 comprising the amino acid sequence of Figure 4 or Figure 5.

13. The apopholasin according to claim 11 or claim 12 when expressed by recombinant DNA or RNA according to claim 10.

14. The apopholasin according to claim 13, which is non-glycosylated.

15. A cell, plasmid, virus or live organism that has been genetically engineered to produce an apoprotein, said cell, plasmid, virus or live organism having incorporated expressibly therein a sequence according to any one of claims 1 to 10.

16. A vector comprising a sequence according to any one of claims 1 to 10.

17. A host cell transformed or transfected with a vector according to claim 16.

18. A BOIP, as defined herein, comprising an apophotoprotein according to any one of claims 11 to 14 in association with a luciferin.

19. A BOIP according to claim 18, wherein the luciferin is derived from *Pholas dactylus*.

20. A method for the preparation of a BOIP, as defined herein, which method comprises bringing an apophotoprotein, such as recombinant apopholasin, into association with a luciferin therefor, such as a luciferin derived from *Pholas dactylus*.

21. A BOIP, apophotoprotein thereof, or a nucleic acid sequence encoding either of these, which comprises a sequence according to any one of Figures 2 to 6 or 9 that has been chemically or genetically modified.

22. A method for the detection and/or measurement of oxygen or one of its metabolites extracellularly, which method comprises providing a BOIP, such as native or chemically- or genetically- modified BOIP or a 'rainbow protein' based on such a BOIP, extracellularly and thereafter detecting and/or quantifying light emission therefrom and/or changes in colour, intensity and/or polarisation of emission(s), wherein the apophotoprotein comprises recombinant apopholasin.

23. A method for the detection and/or measurement of oxygen or one of its metabolites in live cells (intracellularly), which method comprises providing a BOIP, such as native or chemically- or genetically- modified BOIP or a 'rainbow protein' based on such a BOIP, intracellularly and thereafter detecting and/or

quantifying light emission therefrom and/or changes in colour, intensity and/or polarisation of emission(s) therefrom.

24. A method according to claim 22 or 23, wherein said BOIP includes a signal peptide, targetting it to a pre-determined extra- or intra- cellular site.

25. A method according to claim 22 or claim 23, comprising incubating a test sample with a cell according to claim 15 or with a membrane preparation derived therefrom.vnvnvvnbn

26. A method according to any one of claims 22 to 24, wherein light emission takes place in the absence of a luciferase.

27. The use of a sequence or a protein according to any one of claims 1 to 21 in the detection, diagnosis or measurement of oxygen or a metabolite thereof.

28. A diagnostic kit incorporating a sequence or protein according to any one of claims 1 to 21.

29. A process for obtaining a substantially homologous source of apopholasin, which comprises culturing cells having incorporated expressibly therein a polynucleotide encoding apopholasin as defined in any one of claims 1 to 10, and thereafter recovering the cultured cells.

30. A method, use or kit according to any one of claims 20 to 28, substantially as hereinbefore described with particular reference to the Examples.

FIGURE 1

**Clone 40:**

*GAATTCGGCACGAG*TCGGAAAAGAACAAA**ATG**GCTTGTATCGTTTTCGTT
GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAA
GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTGAT
TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC
AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
CTGG**TAG**ATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACC
AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
ATAGAATATTGAAAATAAAATGTTAATAAACACTGGTTGAAATATGAAAA
AAAAAAAAAAAAAAA*CTCGAG*

**Clone 3:**

*GAATTCGGCACGAG*GGAAAAGAACAAA**ATG**GCTTGTATCGTTTTCGTT
*GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA*
ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
*GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAG*
GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
*AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA*
GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
*TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC*
TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
*ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG*
TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTTAT
*TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC*
AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
CTGG**TAG**ATCTATCAGACCACTTTTATCAGCAGGACAACTGGTCGTTACC
AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
ATAGAATATTGAAAATAA

**Clone 5:**

*GT*CGGAAAAGAACAAA**ATG**GCTTGTATCGTTTTCGTTGCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTC
CGGTGAGGAAGTACAATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACTGGATGAC
CATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAGGATCGGGCTTTGGGGCTGTGTCGGATTGA
ACGGGCCGGCCCAGGTACCACAAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAAG
AAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGACTACAGACCACAGGAAGACGGAAC
TGAGAAAACTTTTACAAGAAAATTCTCTAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTAC
AAGGGACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTGTCCCTATATGACGAAGA
TAACAATGGTGTAATGGATGAAGGTAAGGTTATTCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGG
GCTCTTAGACCAAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCAGACACAGTAGA
CGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTTCTGG**TAG**ATCTATCAGACCACTTTTATCAGCAG
GACAACTGGTCGTTACCAGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCGATAGA
ATATTGAAAATAAAATGTTAATAGACACTGGTTGAAAAAAAAAAAAAAAAAAAA*CTCGAG*

FIGURE 2

```
clone 40    GAATTCGGCACGAGTCGGAAAAGAACAAAATGGCTTGTATCGTTTTCGTT
clone 3     GAATTCGGCACGAG--GGAAAAGAACAAAATGGCTTGTATCGTTTTCGTT
clone 5     -------------GTCGGAAAAGAACAAAATGGCTTGTATCGTTTTCGTT
                          *  **********************************

clone 40    GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
clone 3     GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
clone 5     GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
            **************************************************

clone 40    ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
clone 3     ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
clone 5     ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
            **************************************************

clone 40    GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAA
clone 3     GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAG
clone 5     GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAG
            **************************************************

clone 40    GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
clone 3     GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
clone 5     GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
            **************************************************

clone 40    AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
clone 3     AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
clone 5     AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
            **************************************************

clone 40    GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
clone 3     GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
clone 5     GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
            **************************************************

clone 40    TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
clone 3     TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
clone 5     TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
            **************************************************

clone 40    TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
clone 3     TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
clone 5     TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
            **************************************************

clone 40    ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
clone 3     ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
clone 5     ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
            **************************************************

clone 40    TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTGAT
clone 3     TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTTAT
clone 5     TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTTAT
            ********************************************** **

clone 40    TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC
clone 3     TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC
clone 5     TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC
            **************************************************

clone 40    AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
clone 3     AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
clone 5     AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
```

```
                  **************************************************

clone 40          GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
clone 3           GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
clone 5           GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
                  **************************************************

clone 40          CTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACC
clone 3           CTGGTAGATCTATCAGACCACTTTTATCAGCAGGACAACTGGTCGTTACC
clone 5           CTGGTAGATCTATCAGACCACTTTTATCAGCAGGACAACTGGTCGTTACC
                  ****************** ******************************

clone 40          AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
clone 3           AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
clone 5           AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
                  **************************************************

clone 40          ATAGAATATTGAAAATAAAATGTTAATAAACACTGGTTGAAATATGAAAA
clone 3           ATAGAATATTGAAAATAA--------------------------------
clone 5           ATAGAATATTGAAAATAAAATGTTAATAGACACTGGTTGAAA-----AAA
                  ******************

clone 40          AAAAAAAAAAAAAAACTCGAG
clone 3           ---------------------
clone 5           AAAAAAAAAAAAAAACTCGAG
```

**FIGURE 3**

*GAATTCGGCACGAG*TCGGAAAAGAACAAA**ATGGCTTGTATCGTTTTCGT**TGCTCTTG

8S

TCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACAATGCGCGATGAATT

GGACACAAGCTAATGAATATGTGTTCAACGTGGACTGGATGACCATTTTCATCTACG

ACTATGGCGCTCAAGAGCAACTGTACGAAG**ATCGGGCTTTGGGGCTGTGT**CGGATTG

3A

AACGGGCCGGCCCAGGTACCACAAAAGCCGTCTGGATTAACTGGAGTAACGACACGC

AGTCATGTGTAACAAGAAAAACAATCTTCTTCGAGGTTGGTGGAGAAA**TTGCCCGGC**

4S

**TAGTTGAC**TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCT

CTAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGGACGCTG

ATGATAAA**TGCATCGAAGGCACAATTGT**GGTGACAGTCAGGGTGTCCCTATATGACG

6A

AAGATAACAATGGTGTAATGGATGAAGGTAAGGTGTTCCATCTGAGACAATCGAGGA

TGATATCAA**GGACTGTGGGCTCTTAG**ACCAAGATGTTGAACTCGATTATACGTGGAC

7S

TCAAAACGAGTGTGATCTACCAGACACAGTAGACGAGGCTGAAGACACACCGTCAGA

AACTGGAGAATTCTTCTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGG

TCGTTACCAGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATC

GATAGAATATTGAAAATAAAATG**TTAATAAACACTGGTTGAAATATG**AAAAAAAAAA

5A

AAAAAAAA*CTCGAG*

**FIGURE 4A**

Untranslated region
*GAATTC*GGCACGAGTCGGAAAAGAACAAA
Translated region

```
ATG GCT TGT ATC GTT TTC GTT GCT CTT GTC GCT CTA TGC TTA ATG          45
M   A   C   I   V   F   V   A   L   V   A   L   C   L   M

CAA CCG GGT TCC GGT GAG GAA GTA CAA TGC GCG ATG AAT TGG ACA          90
Q   P   G   S   G   E   E   V   Q   C   A   M   N   W   T

CAA GCT AAT GAA TAT GTG TTC AAC GTG GAC TGG ATG ACC ATT TTC         135
Q   A   N   E   Y   V   F   N   V   D   W   M   T   I   F

ATC TAC GAC TAT GGC GCT CAA GAG CAA CTG TAC GAA GAT CGG GCT         180
I   Y   D   Y   G   A   Q   E   Q   L   Y   E   D   R   A

TTG GGG CTG TGT CGG ATT GAA CGG GCC GGC CCA GGT ACC ACA AAA         225
L   G   L   C   R   I   E   R   A   G   P   G   T   T   K

GCC GTC TGG ATT AAC TGG AGT AAC GAC ACG CAG TCA TGT GTA ACA         270
A   V   W   I   N   W   S   N   D   T   Q   S   C   V   T

AGA AAA ACA ATC TTC TTC GAG GTT GGT GGA GAA ATT GCC CGG CTA         315
R   K   T   I   F   F   E   V   G   G   E   I   A   R   L

GTT GAC TAC AGA CCA CAG GAA GAC GGA ACT GAG AAA ACT TTT ACA         360
V   D   Y   R   P   Q   E   D   G   T   E   K   T   F   T

AGA AAA TTC TCT AGC AAA ATG CCA GGC ACT TAC ATG CTT ATG GAC         405
R   K   F   S   S   K   M   P   G   T   Y   M   L   M   D

GTG TGC GCT ACA AGG GAC GCT GAT GAT AAA TGC ATC GAA GGC ACA         450
V   C   A   T   R   D   A   D   D   K   C   I   E   G   T

ATT GTG GTG ACA GTC AGG GTG TCC CTA TAT GAC GAA GAT AAC AAT         495
I   V   V   T   V   R   V   S   L   Y   D   E   D   N   N

GGT GTA ATG GAT GAA GGT AAG GTG ATT CCA TCT GAG ACA ATC GAG         540
G   V   M   D   E   G   K   V   I   P   S   E   T   I   E

GAT GAT ATC AAG GAC TGT GGG CTC TTA GAC CAA GAT GTT GAA CTC         585
D   D   I   K   D   C   G   L   L   D   Q   D   V   E   L

GAT TAT ACG TGG ACT CAA AAC GAG TGT GAT CTA CCA GAC ACA GTA         630
D   Y   T   W   T   Q   N   E   C   D   L   P   D   T   V

GAC GAG GCT GAA GAC ACA CCG TCA GAA ACT GGA GAA TTC TTC TGG         675
D   E   A   E   D   T   P   S   E   T   G   E   F   F   W

TAG ATC TAT CAG ACT ACT TTT ATC AGC AGG ACA ACT GGT CGT TAC         720
*

CAG ACA CCT ATA ACG TGT CCT CAT CAA TAA                             750

* = stop for translation
```

FIGURE 4B

EcoR I

*GAATTCGGCACGAG*TCGGAAAAGAACAAA

**ATG** GCT TGT ATC GTT TTC GTT GCT CTT GTC GCT CTA
TGC TTA ATG CAA CCG GGT TCC GGT GAG GAA GTA CAA
TGC GCG ATG AAT TGG ACA CAA GCT AAT GAA TAT GTG
TTC AAC GTG GAC TGG ATG ACC ATT TTC ATC TAC GAC
TAT GGC GCT CAA GAG CAA CTG TAC GAA GAT CGG GCT
TTG GGG CTG TGT CGG ATT GAA CGG GCC GGC CCA GGT
ACC ACA AAA GCC GTC TGG ATT AAC TGG AGT AAC GAC
ACG CAG TCA TGT GTA ACA AGA AAA ACA ATC TTC TTC
GAG GTT GGT GGA GAA ATT GCC CGG CTA GTT GAC TAC
AGA CCA CAG GAA GAC GGA ACT GAG AAA ACT TTT ACA
AGA AAA TTC TCT AGC AAA ATG CCA GGC ACT TAC ATG
CTT ATG GAC GTG TGC GCT ACA AGG GAC GCT GAT GAT
AAA TGC ATC GAA GGC ACA ATT GTG GTG ACA GTC AGG
GTG TCC CTA TAT GAC GAA GAT AAC AAT GGT GTA ATG
GAT GAA GGT AAG GTG ATT CCA TCT GAG ACA ATC GAG
GAT GAT ATC AAG GAC TGT GGG CTC TTA GAC CAA GAT
GTT GAA CTC GAT TAT ACG TGG ACT CAA AAC GAG TGT
GAT CTA CCA GAC ACA GTA GAC GAG GCT GAA GAC ACA
CCG TCA GAA ACT GGA GAA TTC TTC TGG **TAG**

**ATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACCAGAC**
**ACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCGA**
**TAGAATATTGAAAATAAAATGTTAATAAACACTGGTTGAAATATGAA**
**AAAAAAAAAAAAAAAA***CTCGAG*

Xho I

Figure 4B (bis).

EcoR I

**GAATTC*GGCACGAG*TCGGAAAAGAACAAA**

**ATG** GCT TGT ATC GTT TTC GTT GCT CTT GTC GCT CTA
TGC TTA ATG CAA CCG GGT TCC GGT GAG GAA GTA CAA
TGC GCG ATG AAT TGG ACA CAA GCT AAT GAA TAT GTG
TTC AAC GTG GAC TGG ATG ACC ATT TTC ATC TAC GAC
TAT GGC GCT CAA GAG CAA CTG TAC GAA GAT CGG GCT
TTG GGG CTG TGT CGG ATT GAA CGG GCC GGC CCA GGT
ACC ACA AAA GCC GTC TGG ATT AAC TGG AGT AAC GAC
ACG CAG TCA TGT GTA ACA AGA AAA ACA ATC TTC TTC
GAG GTT GGT GGA GAA ATT GCC CGG CTA GTT GAC TAC
AGA CCA CAG GAA GAC GGA ACT GAG AAA ACT TTT ACA
AGA AAA TTC TCT AGC AAA ATG CCA GGC ACT TAC ATG
CTT ATG GAC GTG TGC GCT ACA AGG GAC GCT GAT GAT
AAA TGC ATC GAA GGC ACA ATT GTG GTG ACA GTC AGG
GTG TCC CTA TAT GAC GAA GAT AAC AAT GGT GTA ATG
GAT GAA GGT AAG GTG ATT CCA TCT GAG ACA ATC GAG
GAT GAT ATC AAG GAC TGT GGG CTC TTA GAC CAA GAT
GTT GAA CTC GAT TAT ACG TGG ACT CAA AAC GAG TGT
GAT CTA CCA GAC ACA GTA GAC GAG GCT GAA GAC ACA
CCG TCA GAA ACT GGA GAA TTC TTC TGG **TAG**

**ATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACCAGAC**
**ACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCGA**
**TAGAATATTGAAAATAAAATGTTAATAAACACTGGTTGAAATATGAA**
**AAAAAAAAAAAAAAAACTCGAG**

Xho I

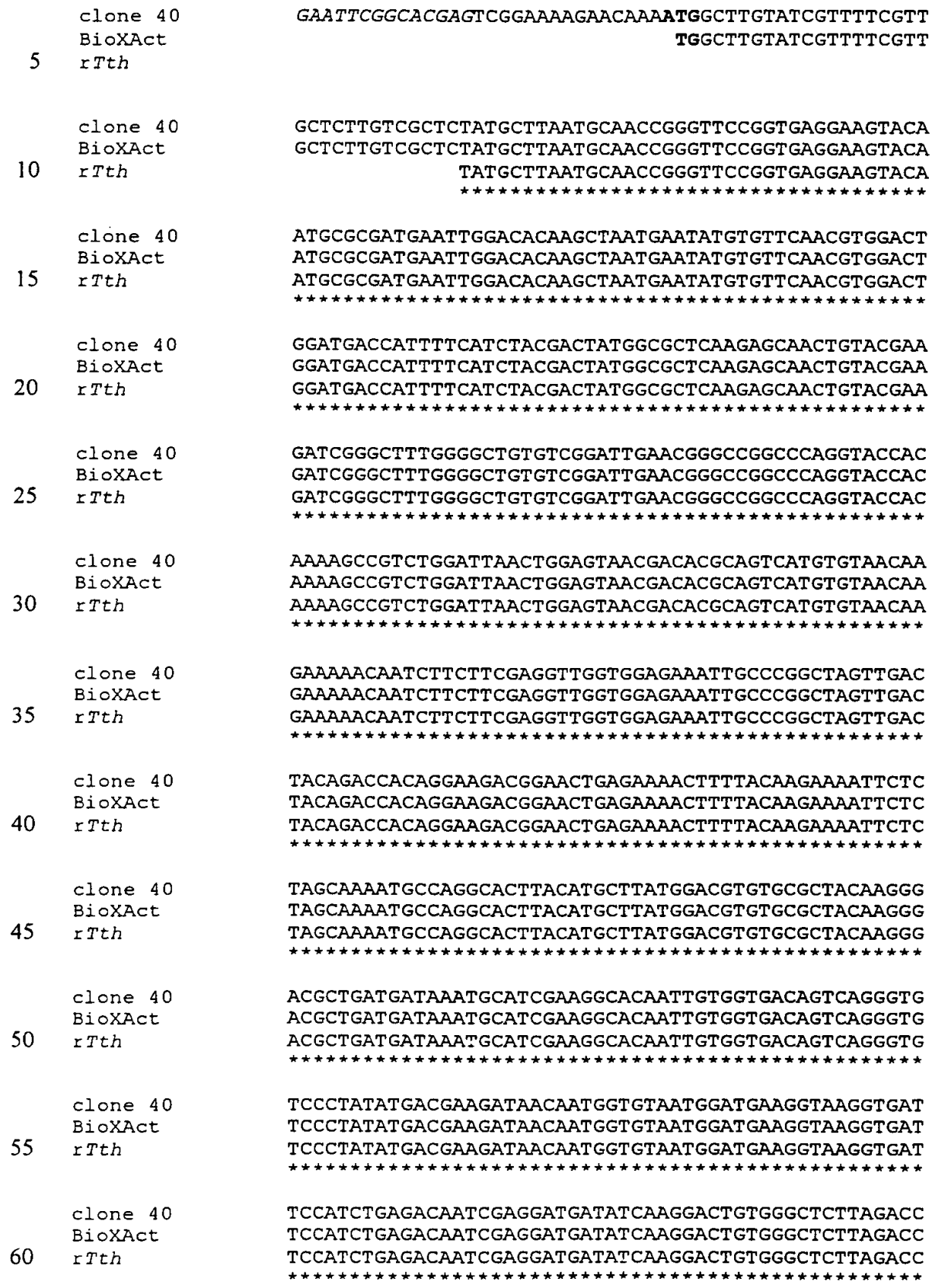
**FIGURE 5A**

EEVQCAMNWTQANEYVFNVDWMTIFIYDYGAQEQLYEDRALGLCRIERAGPGTTKAVWIN
WSNDTQSCVTRKTIFFEVGGEIARLVDYRPQEDGTEKTFTRKFSSKMPGTYMLMDVCATR
DADDKCIEGTIVVTVRVSLYDEDNNGVMDEGKVIPSETIEDDIKDCGLLDQDVELDYTWT
QNECDLPDTVDEAEDTPSETGEFFW

**FIGURE 5B**

MACIVFVALVALCLMQPGSGEEVQCAMNWTQANEYVFNVDWMTIFIYDYGAQEQLYEDRA
LGLCRIERAGPGTTKAVWINWSNDTQSCVTRKTIFFEVGGEIARLVDYRPQEDGTEKTFT
RKFSSKMPGTYMLMDVCATRDADDKCIEGTIVVTVRVSLYDEDNNGVMDEGKVIPSETIE
DDIKDCGLLDQDVELDYTWTQNECDLPDTVDEAEDTPSETGEFFW

**FIGURE 6**


clone 40 GAATTCGGCACGAGTCGGAAAAGAACAAAATGGCTTGTATCGTTTTCGTT
BioXAct TGGCTTGTATCGTTTTCGTT
rTth
clone 40 GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
BioXAct GCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
rTth TATGCTTAATGCAACCGGGTTCCGGTGAGGAAGTACA
clone 40 ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
BioXAct ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
rTth ATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGACT
clone 40 GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAA
BioXAct GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAA
rTth GGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAA
clone 40 GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
BioXAct GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
rTth GATCGGGCTTTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCAC
clone 40 AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
BioXAct AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
rTth AAAAGCCGTCTGGATTAACTGGAGTAACGACACGCAGTCATGTGTAACAA
clone 40 GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
BioXAct GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
rTth GAAAAACAATCTTCTTCGAGGTTGGTGGAGAAATTGCCCGGCTAGTTGAC
clone 40 TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
BioXAct TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
rTth TACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACAAGAAAATTCTC
clone 40 TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
BioXAct TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
rTth TAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGGG
clone 40 ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
BioXAct ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
rTth ACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTG
clone 40 TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTGAT
BioXAct TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTGAT
rTth TCCCTATATGACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTGAT
clone 40 TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC
BioXAct TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC
rTth TCCATCTGAGACAATCGAGGATGATATCAAGGACTGTGGGCTCTTAGACC

```
clone 40        AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
BioXAct         AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
rTth            AAGATGTTGAACTCGATTATACGTGGACTCAAAACGAGTGTGATCTACCA
                **************************************************

clone 40        GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
BioXAct         GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
rTth            GACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACTGGAGAATTCTT
                **************************************************

clone 40        CTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACC
BioXAct         CTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACC
rTth            CTGGTANATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTACC
                ****** **************************************************

clone 40        AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
BioXAct         AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAACAGAAATAATCG
rTth            AGACACCTATAACGTGTCCTCATCAATAATGTGTAAAAC
                *************************************

clone 40        ATAGAATATTGAAAATAAAATGTTAATAAACACTGGTTGAAATATGAAAA
BioXAct         ATAGAATATTGAAAATAAAATGTTAATAAACACTGGTTGAAATATGAA
rTth

clone 40        AAAAAAAAAAAAAACTCGAG
BioXAct
rTth
```

**FIGURE 7A**

Oligo 1

ACI ATH TTY TTY CAR GT

Oligo 2

CAR GAR GAR GGN ACI GA

Oligo 2A

TCI GTN CCY TCY TCY TG

Oligo N

TTY AAY GTI GAY TGG ATG

| | | | | | |
|---|---|---|---|---|---|
| M=A/C | R=A/G | W=A/T | S=G/C | Y=C/T | K=G/T |
| V=A/C/G | H=A/C/T | D=A/G/T | B=C/G/T | N=A/C/G/T | I=inosine |

**FIGURE 7B**

Oligo 3A

ACA CAG CCC CAA AGC CCG AT

Oligo 4S

TTG CCC GGC TAG TTG ACT AC

Oligo 5A

CAT ATT TCA ACC AGT GTT TAT TAA

Oligo 6A

CAA TTG TGC CTT CGA TGC A

Oligo 7S

GGA CTG TGG GCT CTT AG

Oligo 8S

ATG GCT TGT ATC GTT TTC GT

Oligo T7

**FIGURE 7C**

Oligo ExS

CCA CAC GGA TCC TGA GGA AGT ACA ATG

Oligo ExA

CCA CAC GGA TCC TTA TTG ATG AGG ACA

Oligo Bac1

CTT GTT TTT ATG GTC GTC TAC ATT TCT TAC ATC TAT GCG GAG GAA GTA CAA TG

Oligo C9 12

CCA CAC AGA TCT AGA ATG AAA TTC TTA GTC AAC GTT GCC CTT GTT TTT ATG GTC

***Oligo BV5***

TTT ACT GTT TTC GTA ACA GTT TTG

***Oligo BV3***

CAA CAA CGC ACA GAA TCT AG

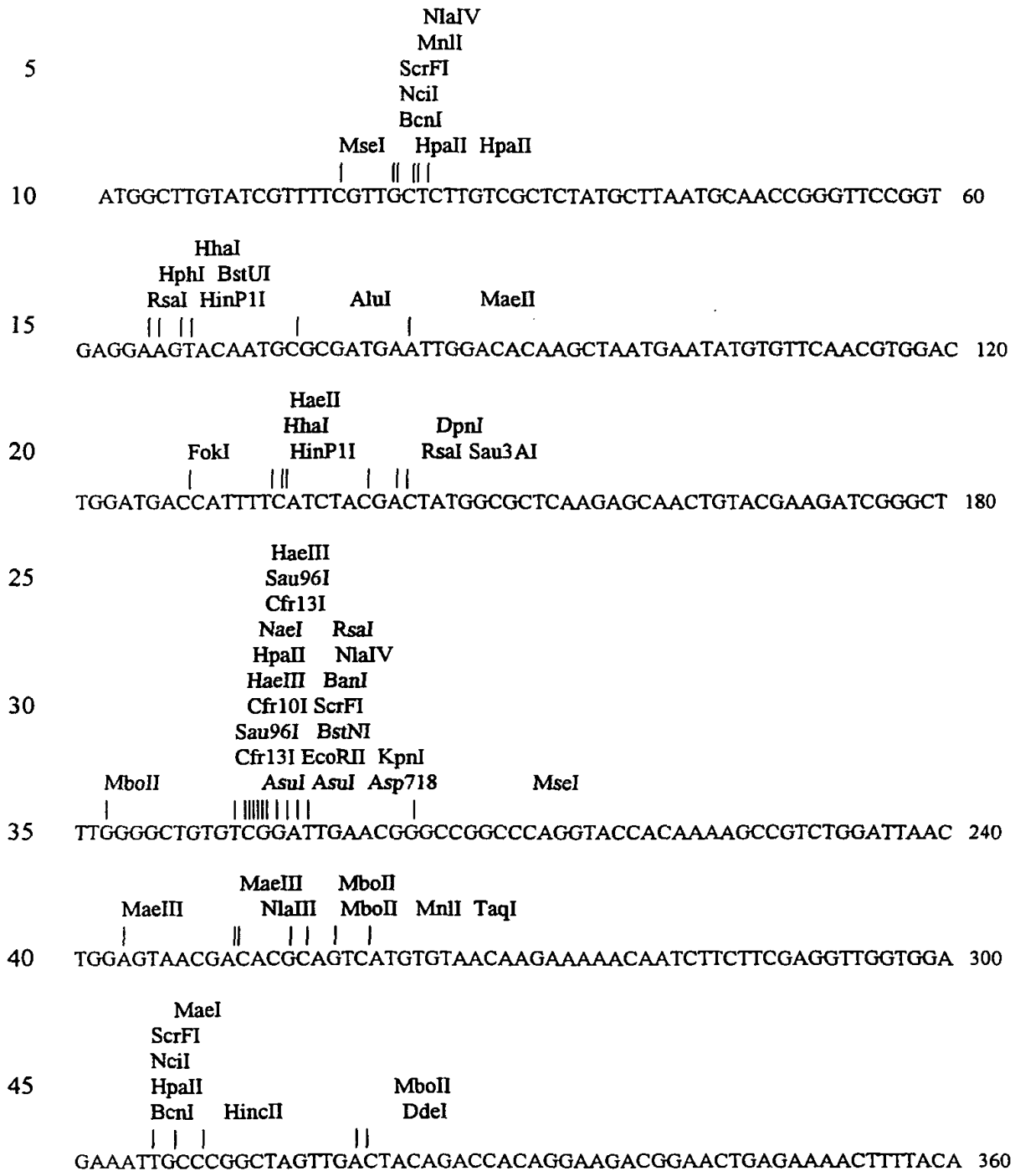
FIGURE 8

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AccI | 630 | | | | | | |
| | AflIII | 405 | 734 | | | | | |
| 5 | AluI | 95 | | | | | | |
| | AlwNI | 659 | | | | | | |
| | Asp | 718 | 215 | | | | | |
| | AsuI | 204 | 209 | | | | | |
| | BanI | 215 | | | | | | |
| 10 | BanII | 564 | | | | | | |
| | BcnI | 51 | 310 | | | | | |
| | BglII | 678 | | | | | | |
| | Bsp1 | 286 | 564 | | | | | |
| | BstNI | 213 | 384 | | | | | |
| 15 | BstUI | 77 | | | | | | |
| | Cfr10I | 206 | | | | | | |
| | Cfr13I | 204 | 209 | | | | | |
| | DdeI | 345 | 528 | 565 | | | | |
| | DpnI | 174 | 615 | 680 | | | | |
| 20 | EcoRI | 665 | | | | | | |
| | EcoRII | 211 | 382 | | | | | |
| | EcoRV | 547 | | | | | | |
| | FokI | 136 | 518 | 554 | | | | |
| | HaeII | 153 | | | | | | |
| 25 | HaeIII | 206 | 210 | | | | | |
| | HgaI | 431 | | | | | | |
| | HhaI | 77 | 152 | 413 | | | | |
| | HincII | 319 | | | | | | |
| | HinfI | 520 | 598 | | | | | |
| 30 | HinP1I | 75 | 150 | 411 | | | | |
| | HpaII | 50 | 57 | 207 | 310 | | | |
| | HphI | 71 | 469 | 529 | | | | |
| | KpnI | 219 | | | | | | |
| | MaeI | 314 | 372 | | | | | |
| 35 | MaeII | 114 | 405 | 593 | 734 | | | |
| | MaeIII | 245 | 265 | 457 | 716 | | | |
| | MboII | 182 | 274 | 277 | 347 | 497 | 653 | 661 |
| | MnlI | 54 | 282 | 531 | 627 | 750 | | |
| | MseI | 41 | 237 | | | | | |
| 40 | NaeI | 208 | | | | | | |
| | NciI | 51 | 310 | | | | | |
| | NlaIII | 264 | 397 | | | | | |
| | NlaIV | 55 | 217 | | | | | |
| | NsiI | 440 | | | | | | |
| 45 | NspHI | 397 | | | | | | |
| | PleI | 592 | | | | | | |
| | RsaI | 69 | 167 | 217 | | | | |
| | Sau3AI | 172 | 613 | 678 | | | | |

12/16

| | | | | |
|---|---|---|---|---|
| Sau96I | 204 | 209 | | |
| ScrFI | 51 | 213 | 310 | 384 |
| SfaNI | 428 | | | |
| TaqI | 288 | 441 | 537 | 585 |
| XhoII | 678 | | | |

**FIGURE 9**


NlaIV
MnlI
ScrFI
NciI
BcnI
MseI HpaII HpaII
ATGGCTTGTATCGTTTTCGTTGCTCTTGTCGCTCTATGCTTAATGCAACCGGGTTCCGGT 60
HhaI
HphI BstUI
RsaI HinP1I AluI MaeII
GAGGAAGTACAATGCGCGATGAATTGGACACAAGCTAATGAATATGTGTTCAACGTGGAC 120
HaeII
HhaI DpnI
FokI HinP1I RsaI Sau3AI
TGGATGACCATTTTCATCTACGACTATGGCGCTCAAGAGCAACTGTACGAAGATCGGGCT 180
HaeIII
Sau96I
Cfr13I
NaeI RsaI
HpaII NlaIV
HaeIII BanI
Cfr10I ScrFI
Sau96I BstNI
Cfr13I EcoRII KpnI
MboII AsuI AsuI Asp718 MseI
TTGGGGCTGTGTCGGATTGAACGGGCCGGCCCAGGTACCACAAAAGCCGTCTGGATTAAC 240
MaeIII MboII
MaeIII NlaIII MboII MnlI TaqI
TGGAGTAACGACACGCAGTCATGTGTAACAAGAAAAACAATCTTCTTCGAGGTTGGTGGA 300
MaeI
ScrFI
NciI
HpaII MboII
BcnI HincII DdeI
GAAATTGCCCGGCTAGTTGACTACAGACCACAGGAAGACGGAACTGAGAAAACTTTTACA 360

```
                ScrFI
                BstNI      NspHI  MaeII HinP1I
          MaeI  EcoRII       NlaIII AflIII HhaI
          |     ||     |     |    ||
AGAAAATTCTCTAGCAAAATGCCAGGCACTTACATGCTTATGGACGTGTGCGCTACAAGG 420

          HgaI  TaqI
          SfaNI NsiI        MaeIII   HphI
          | |   ||          |        |
GACGCTGATGATAAATGCATCGAAGGCACAATTGTGGTGACAGTCAGGGTGTCCCTATAT 480

                                  MnlI
                            HinfI HphI
            MboII           FokI  DdeI  TaqI
            |               ||    |||   |
GACGAAGATAACAATGGTGTAATGGATGAAGGTAAGGTGATTCCATCTGAGACAATCGAG 540

                  DdeI
                  Bsp1286             MaeII
      EcoRV FokI  BanII               TaqI  PleI HinfI
      |   |   ||                      |   ||  |
GATGATATCAAGGACTGTGGGCTCTTAGACCAAGATGTTGAACTCGATTATACGTGGACT 600

           DpnI       AccI
           Sau3AI     MnlI                MboII AlwNI
           ||         | |                 |    |
CAAAACGAGTGTGATCTACCAGACACAGTAGACGAGGCTGAAGACACACCGTCAGAAACT 660

             DpnI
             XhoII
     EcoRI   Sau3AI
  MboII      BglII                    MaeIII
  | |        ||                       |
GGAGAATTCTTCTGGTAGATCTATCAGACTACTTTTATCAGCAGGACAACTGGTCGTTAC 720

           MaeII
           AflIII     MnlI
           |          |
CAGACACCTATAACGTGTCCTCATCAATAA                              750
```

Pc[redacted] NO · GB99/0[redacted]54

Form 23/77 · 19.11.99

Agent · Wynne-Jones, Laird & James